

UNITED STATES DISTRICT COURT
DISTRICT OF MINNESOTA

CHRISTOPHER KOHLS and
MARY FRANSON,

Court File No. 24-cv-03754 (LMP/DLM)

Plaintiffs,

vs.

**EXPERT DECLARATION OF
PROFESSOR JEVIN WEST**

KEITH ELLISON, in his official capacity
as Attorney General of Minnesota, and
CHAD LARSON, in his official capacity
as County Attorney of Douglas County.

Defendants.

I, Jevin West, state as follows:

1. I am a Professor in the Information School at the University of Washington and the co-founder and inaugural director of the Center for an Informed Public, an interdisciplinary research center dedicated to studying misinformation in the digital age.

2. For nearly ten years, my research and teaching have focused on the study of misinformation¹. I wrote a book² and developed a class³ on misinformation that has been adopted at more than 100 universities. I co-founded and was the inaugural director of the Center for an Informed Public⁴ (CIP) at the University of Washington (UW). It is an internationally recognized center focused on understanding the spread of strategic

¹ <https://jevinwest.org/>

² <https://www.penguinrandomhouse.com/books/563882/calling-bullshit-by-carl-t-bergstrom-and-jevin-d-west/>

³ <https://callingbullshit.org/>

⁴ <https://cip.uw.edu/>

misinformation. In my research career, I have published more than 100 papers in journals, books and conference proceedings in many of the top journals including the *Proceedings of the Academy of Sciences*, *Nature Human Behavior*, *Science Advances*, etc; I have also published patents, op-eds, blog posts, and white papers. I have given more than 300 invited lectures around the globe, including over 40 keynotes at major conferences and universities. My work has been featured or I have been quoted in over 450 news articles, including the *Wall Street Journal*, the *New York Times*, the *Economist*, *Wired*, the *Washington Post*, *Nature*, *Science*, and many others.

My core research focuses on understanding and mitigating the spread of misinformation and its effects on society. I will provide four examples to illustrate the type of research and work that I do.

- a. In the CIP, my colleagues and I received a National Science Foundation grant⁵ to study and develop rapid response research methods for mitigating the spread of online disinformation, especially during major world events, crises, and elections. Instead of waiting five years for a research article to be published on the spreading dynamics of election rumors, our goal is to make our results available to the public in a matter of days. Our work includes analyses about the U.S. elections⁶ and recent events⁷ surrounding the hurricane in North Carolina⁸. We have examined the ways that social media, traditional media, synthetic media (e.g., deepfakes), and influencers embedded in these media systems amplify the spread of misinformation.
- b. I am one of the co-leads for a forthcoming *National Academy of Sciences, Engineering, and Medicine* Consensus Report on misinformation about science. The multi-year effort has examined the vast literature that has grown around misinformation, its origins, its effects on society, and the interventions to slow

⁵ https://www.nsf.gov/awardsearch/showAward?AWD_ID=2120496

⁶ <https://theconversation.com/5-types-of-misinformation-to-watch-out-for-while-ballots-are-being-counted-and-after-149509>

⁷ <https://www.cip.uw.edu/2024/09/05/what-to-expect-when-were-electing-2024/>

⁸ <https://www.cip.uw.edu/2024/10/09/hurricane-helene-election-rumors/>

its spread. The goal of the report is to establish what is known and not known about misinformation, specifically misinformation in and about science, and to make this work accessible to educators, policymakers, journalists, and researchers outside the field of misinformation studies.

- c. Many of my recent peer-reviewed publications have focused on auditing social media platforms and search engines, studying the factors that lead to rumor spreading, and examining the different types of interventions employed. For example, I have published papers that have audited Google's search engine during the 2020 US election⁹, examined the impact of recommender systems on social media on forming echo chambers¹⁰, and analyzed various intervention strategies¹¹, looking both at the pros and cons, for mitigating the spread of misinformation on social media.
- d. I have spent considerable time developing education curriculum and helping legislators think about strategies for dealing with misinformation. For example, I developed one of the first games to bring public attention to style-GAN (Generative Adversarial Network) images, one of the precursors to modern deepfakes. The game is called Whichfaceisreal.com¹² and has been played tens of millions of times around the world. The game's goal was simple. Players were asked to identify the human photo that was real. It turns out it was not so easy in 2019, and it has only become more difficult over time. I also have created online quizzes¹³ and handbooks¹⁴ about deepfakes in collaboration with Microsoft to further prepare the public for deepfakes. In addition to games, I co-created the Misinfoday¹⁵ program, which has brought thousands of students to college campus each year to learn about media literacy and things like deepfakes. The program has expanded from the state of Washington to other states in the U.S. and even other countries. The program has also expanded into libraries and community organizations focused on adult education. In addition to education efforts, I have helped legislators write legislation around deepfakes. This

⁹ Zade, H., Wack, M., Zhang, Y., Starbird, K., Calo, R., Young, J. and West, J.D., 2022. Auditing Google's Search Headlines as a Potential Gateway to Misleading Content: Evidence from the 2020 US Election. *Journal of Online Trust and Safety*, 1(4).

¹⁰ Duskin, Kayla, et al. "Echo Chambers in the Age of Algorithms: An Audit of Twitter's Friend Recommender System." *Proceedings of the 16th ACM Web Science Conference*. 2024.

¹¹ Bak-Coleman, Joseph B., et al. "Combining interventions to reduce the spread of viral misinformation." *Nature Human Behaviour* 6.10 (2022): 1372-1380.

¹² <https://www.whichfaceisreal.com/>

¹³ <https://www.spotdeepfakes.org/en-US>

¹⁴ https://jevinwest.org/papers/Prochaska2020CIP_Deepfake_Report.pdf

¹⁵ <https://www.cip.uw.edu/misinfoday/>

includes Washington States Senate Bill 5152, which addresses the use of political deepfakes before an election¹⁶.

3. A full overview of my professional experience and publications is provided in my curriculum vitae, which is attached as **Exhibit A**.

4. I have further identified the academic, scientific, and other materials referenced in this declaration in the references attached as **Exhibit B**.

5. I have been retained by the Office of the Minnesota Attorney General to provide expert opinion and testimony on deepfakes and synthetic media, the spread of disinformation, and the risks that deepfakes pose to elections and democracy. I have reviewed Minnesota's deepfake law, Minnesota Statutes section 609.771, and the Complaint in this case. I am being compensated at the rate of \$500 per hour.

6. Based on my training and professional and experience, I provide my expert views, based on my training and professional experience, on the following topics:

- the rise and rapid advancement of “synthetic” media, including deepfakes, and how technology enables the production of deepfakes;
- the dissemination of deepfakes, including the role of social media as the primary vectors of deepfakes and other mis- and disinformation;
- the risks that deepfakes pose to elections; and
- the challenges of combatting deepfakes, including how the visceral, seemingly authentic nature of deepfakes impacts the effectiveness of traditional fact-checking and debunking efforts.

¹⁶ <https://www.cip.uw.edu/2023/06/09/new-wa-law-deepfake-disclosure-election-media/>

I. DEEPFAKES AND SYNTHETIC MEDIA: DEFINITIONS AND TECHNOLOGY.

7. Deepfakes are images, audio, or video that mimic real and nonexistent people saying and doing things that never happened¹⁷. Some of the core technology actually originated by my colleagues at my university (University of Washington) when they wrote a paper demonstrating how to make Obama say whatever they wanted¹⁸. The term, ‘deepfake’ reflects its fakeness and its underlying technology, ‘deep learning’, which is a method in the field of Artificial Intelligence (AI).

8. The novelty of deepfakes is more than just its fakery. Software tools and older versions of AI existed prior to deepfakes. For example, an adept graphic designer or movie production company 15 years ago could create images and videos of presidents doing and saying things they never did. However, the resources and training required to create convincing videos at the level of current deepfakes was high and extensive. It required professionals with experience in visual effects and computer-generated imagery and access to expensive video editing software. The personnel and software costs could be in the tens of thousands to even millions of dollars, depending on the video, and it would take months to create.

9. The novelty of deepfakes is the ability of the everyday person to now create photorealistic images in a matter of minutes at very little cost. There are even how-to guides

¹⁷ Westerlund, Mika. "The emergence of deepfake technology: A review." *Technology innovation management review* 9.11 (2019).

¹⁸ Suwajanakorn, Supasorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. "Synthesizing obama: learning lip sync from audio." *ACM Transactions on Graphics (ToG)* 36.4 (2017): 1-13.

for how to make them in high profile journalistic venues¹⁹. Prior to deepfakes, it would take an expert designer days, if not weeks and months, to create convincing images and video. Now it is as simple as submitting a prompt and paying a small subscription fee to the many companies offering this service. The scale and speed are the novelties and the reasons why these technologies pose challenges to the integrity of our information environments.

10. Synthetic media is a broader term that includes deepfakes but also text-generated content, AI art, music composition, etc. All deepfakes are considered a form of synthetic media, but not all synthetic media is a deepfake. Synthetic media, as a concept, has been around longer than deepfakes. While deepfakes are typically viewed negatively, synthetic media is viewed more neutrally. There are good and bad uses of synthetic media. Deepfakes, on the other hand, tend to be viewed more negatively because of the way that they are employed and strategically distributed in a political settings. Researchers have shown that deepfakes, especially those microtargeted to specific groups, can impact people's attitudes towards a politician²⁰ and can impact trust in news²¹.

11. Although there are many techniques employed to construct deepfakes (e.g., face-swapping algorithms, recurrent neural networks, voice cloning, autoencoders, and convolutional neural networks, etc.), Generative Adversarial Networks²² (GANs) are the

¹⁹ <https://www.pnas.org/doi/10.1073/pnas.2315678121>

²⁰ Dobber, Tom, et al. "Do (microtargeted) deepfakes have real effects on political attitudes?." *The International Journal of Press/Politics* 26.1 (2021): 69-91.

²¹ Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." *Social media+ society* 6.1 (2020): 2056305120903408.

²² Goodfellow, Ian, et al. "Generative adversarial networks." *Communications of the ACM* 63.11 (2020): 139-144.

core technology. With this approach, two computer programs are pitted against each other—one attempting to create new content that cannot be detected as fake, the generator, and another detecting whether the created image is real, the discriminator. This general approach is one of the reasons why detecting deepfakes is so difficult. As soon as the detector achieves some level of proficiency in detecting a deepfake, the image creator figures out something new to fool the detector.

12. Deepfake technology is changing rapidly. When deepfakes first hit the scene, they were good, but they had weaknesses. For example, deepfake images struggled with human hair, teeth, earrings, symmetry of glasses, etc²³. They also could not produce the same deepfake face in multiple positions. These weaknesses have nearly been erased. Fake images of human faces can be shown in multiple angles, hair has improved, etc. In addition, the time it takes to create a deepfake image, video or audio file is a fraction of the time it used to take, even just a couple years ago. One can now create a deepfake audio with less than a minute of sample audio of the person being mimicked²⁴, and they have become exceedingly hard to detect²⁵ in many different settings²⁶ as recent research has shown²⁷.

²³ <https://www.whichfaceisreal.com/learn.html>

²⁴ <https://www.theverge.com/2017/4/24/15406882/ai-voice-synthesis-copy-human-speech-lyrebird>

²⁵ <https://www.wired.com/story/generative-ai-detection-gap/>

²⁶ <https://www.wired.com/story/generative-ai-detection-gap/>

²⁷ <https://www.nature.com/articles/s41467-024-51998-z>

II. THE DISSEMINATION OF DEEPPAKES AND OTHER MIS- AND DISINFORMATION

13. Deepfakes are found everywhere on the internet. They spread most effectively, as does most information on the internet, through social media platforms. This includes Facebook, Instagram, X, YouTube, TikTok, Reddit, 4chan, Telegram, WhatsApp, and many others. But they can come through other mediums as well. For example, an audio deepfake was used to push messages of President Biden encouraging Democrats to not vote in the 2024 New Hampshire primary²⁸. Another common example involves scams over phone and email. Attorney Gary Schildhorn testified to the U.S. Senate last last year (2023) about how he nearly fell victim to a scam in which the cloned voice of his son was asking for a wire transfer of \$9,000 for bail money²⁹.

14. My colleagues and I have observed the ways in which disinformation spreads, not just from state actors, but also from ordinary citizens both knowingly and unknowingly³⁰. Participatory disinformation describes this phenomenon. It refers to the spread of false information by ordinary individuals that are often not aware of the larger narrative. Rather than passive consumers of disinformation campaigns, citizens actively engage in the creation and amplification of this content. This bottom-up version of disinformation contrasts with the bottom-down version where one actor or organization

²⁸ <https://www.theguardian.com/technology/article/2024/aug/22/fake-biden-robocalls-fine-lingo-telecom>

²⁹ <https://www.kiro7.com/news/local/father-warns-congress-about-ai-scammer-who-sounded-just-like-his-son/KA7BXJJ2OJB3NHDDM4EGB5L24M/>

³⁰ Prochaska, Stephen, et al. "Mobilizing manufactured reality: How participatory disinformation shaped deep stories to catalyze action during the 2020 US presidential election." *Proceedings of the ACM on human-computer interaction* 7.CSCW1 (2023): 1-39.

drives the disinformation. Participatory disinformation can be more effective when done through participatory mechanisms because of how it is embedded more deeply in the public conversation. It is more difficult to detect its origins. Participatory disinformation often involves deepfakes. Everyday users are active participators in deepfake distribution. They share and comment on the videos as if they are real. They sometimes do this genuinely believing that the deepfake is real. Other times they do it knowing it is not real. Users can adapt a video or image for a given community or context. This is what makes it so difficult for fact checkers and journalists trying to correct the record. It becomes embedded so quickly with buy in from the community, who are contributors and sharers themselves, that it is difficult to debunk and slow.

15. Cross platform spread is a challenge to contain, not just for deepfakes, but for any kind of content, especially disinformation campaigns³¹. You will often see content posted on multiple platforms at once. This is done to garner the biggest audience possible. Even if one platform can stop the spread of something verifiably false, another platform may have no such rules or no desire to stop its spread. The same variation of policy around deepfakes exists. New policy is beginning to emerge at the federal level in the U.S. to stop at least the spread of certain kinds of deepfakes across platforms. For example, the Deceptive Act, which passed in the U.S. Senate with bi-partisan support early this year

³¹ Wilson, Tom, and Kate Starbird. "Cross-platform disinformation campaigns: lessons learned and next steps." *Harvard Kennedy School Misinformation Review* 1.1 (2020).

(2024), would allow victims to sue creators of deepfakes that depict nonconsensual sexually explicit videos³². This now moves to the House.

16. Deepfakes generally are designed to go viral. Deepfake creators post their videos of politicians saying certain things, not because they want to see it in their web browser, but because they want it to be seen by many. Deepfakes are the perfect fodder for attention-grabbing content. They are visually and auditorily engaging; they are plausibly realistic, so much so that users often cannot tell whether they are real or not; and, importantly, they often carry shock value that fits within a larger cultural narrative, of something that many think “could happen.” These are the ingredients for viral sharing and spreading on social media.

III. THE RISKS OF DEEPFAKES TO ELECTIONS AND DEMOCRACY.

17. By design, deepfakes are convincing. They mimic real people saying and doing things that are plausible. In a political context, they often convey a scenario that aligns with the prevailing narrative. For example, if a politician accuses their opponent of being soft on crime, you might expect to see a deepfake of the opponent “saying” that they want to release all low offending prisoners. Because of the relatively low cost of creating deepfakes, one can make more than one deepfake heard by all. Often more convincingly, one could make thousands of customized deepfakes for different communities. Because of the scalability of deepfakes, it can be difficult to fact check deepfakes before they go viral within a community. Because of these reasons, deepfakes can be tools to be used at “the

³² <https://www.nbcnews.com/tech/tech-news/defiance-act-passes-senate-allow-deepfake-victims-sue-rcna163464>

eleventh hour” (see examples below). And even when they fact checked, deepfakes can leave a lasting effect that deepens distrust in our information systems³³.

18. The Liar’s Dividend describes the scenario when someone denies a real event caught in a photograph, video, or audio file, using the following false claim: “that is not real; it is a deepfake.” Because we now live in a world where deepfakes exist, this argument is now possible. It is difficult, if not impossible in some scenarios, to tell whether a piece of content is real or synthetically created. The Liar’s Dividend may be as problematic as deepfakes themselves. As we proliferate our information systems with more and more synthetic, but realistically-looking, content, the more useful and potentially dangerous the Liar’s Dividend.

19. Deepfakes have been distributed in elections around the globe³⁴. They have also been seen throughout the U.S. election cycle this year³⁵. They are often introduced at the “eleventh hour”, or right before the election (days or even hours, rather than weeks) before an election. We saw this for the recent Slovakian election³⁶. This is problematic because it reduces the opportunities to fact check and label the content as fake. And even when it is identified as deepfake and labeled as a deepfake, many of the original users that

³³ Hancock, Jeffrey T., and Jeremy N. Bailenson. "The social impact of deepfakes." *Cyberpsychology, behavior, and social networking* 24.3 (2021): 149-152.

³⁴ <https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/>

³⁵ <https://nymag.com/intelligencer/article/how-ai-deepfakes-influence-disrupt-trump-biden-2024-election.html>

³⁶ <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>

saw the original video may never see the correction. This a general challenge, but especially so the closer to the election. I will provide four recent examples, but there are many others.

- a. In the 2024 Slovakian election, a deepfake emerged on social media a few days before citizens went to the polls³⁷. The fake audio file was of one of the candidates saying they rigged the election; it followed a video deepfake in which the same candidate said he planned to raise the price of beer. The candidate never said any of this, but the deepfakes went viral. That candidate, in that close election, lost. It would be difficult, if not impossible, to show that this video altered the election, just as it is for any causal analysis of large events like an election, but that is not reason to ignore the potential impact that could have on an election.
- b. On the day before the New Hampshire primary in 2024, a deepfake audio call was sent to thousands of people in New Hampshire³⁸. The call sounded like President Biden discouraging people from voting. The Federal Trade Commission proposed fining Steve Kramer, the originator of the deepfake call, \$6 million.
- c. On the day of a Taiwanese election, an affiliated group with the Chinese Communist Party posted a deepfake of one of the main candidates endorsing another candidate, which the politician never did³⁹. Youtube eventually pulled the video, but like many of these videos, it had been seen by many by the time it was pulled.
- d. On the day before the Chicago mayoral race in 2023, a deepfake of Paul Vallas, one of the candidates of the race, emerged⁴⁰. It was eventually taken down, but not before it had been viewed by thousands of people.

³⁷ <https://www.cnn.com/2024/02/01/politics/election-deepfake-threats-invs/index.html>

³⁸ <https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative>

³⁹ <https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/>

⁴⁰ <https://www.cbsnews.com/chicago/news/vallas-campaign-deepfake-video/>

IV. WHY DEEPPAKES ARE UNIQUELY DIFFICULT TO COMBAT WITH TRADITIONAL INTERVENTIONS.

20. Deepfakes are especially challenging to address because they are so easy to make, easy to distribute, and hyper-realistic so it takes time to do the digital forensics to determine whether it is real or not. By the time a fact-checker debunks one deepfake in one community, there are already five more spreading virally on social media. Deepfake detection tools exist⁴¹. Some are better than others⁴², and some are available from for-profit companies and some from non-profit groups. However, none of them are perfect and the deepfakes themselves are only getting better. I am on the board of TrueMedia.org⁴³, which is a non-profit that provides deepfake detection tools to journalists, researchers, and the public. As mentioned above, the tools are far from perfect and probabilistic at best. They can predict, say with 80% confidence, that the video is fake, but many times the results are inconclusive. It then takes hours and hours of manual work, and sometimes days, to sort out whether it is a real video. And many times, those manual efforts are inconclusive. The detectors will get better but so will the technology to create the deepfakes. I do not anticipate a time when a detector is even close to full proof. In the best-case scenario, we will have some tools to help the fact checkers and platforms at least triage suspicious synthetic content.

⁴¹ Rana, Md Shohel, et al. "Deepfake detection: A systematic literature review." *IEEE access* 10 (2022): 25494-25513.

⁴² Dolhansky, Brian, et al. "The deepfake detection challenge (dfdc) dataset." *arXiv preprint arXiv:2006.07397* (2020).

⁴³ <https://www.truemedia.org/>

21. Social media platforms are designed to grab our attention. Once the algorithms identify what grabs the attention of its users, it is built to quickly amplify that content, and it does so often without regard to veracity. As I note above, deepfakes are the perfect fodder for attention-grabbing content. They are visually and auditorily engaging; they are plausibly realistic, so much so that users often cannot tell whether they are real or not; and, importantly, they often carry shock value that fits within a larger cultural narrative, of something that many believe “could happen.” These are the ingredients for viral sharing and spreading on social media.

22. It is difficult to counter what people see and hear with their own eyes and ears, especially when an image or video confirms our biases or narratives of how the world works. And even when an image, video or audio clip depicts something that really happened, authoritative figures can deny their existence with the Liar’s Dividend, as noted above. Deepfakes can erode confidence and trust in our information systems⁴⁴ and ultimately our institutions and authoritative figures. Traditional authorities are having a difficult time being heard, even when deepfakes are not a part of the conversation. Add deepfakes into the equation and countering falsehoods becomes an almost Sisyphean task. But if we care about democracy and an informed citizenry, we must at least make the public aware of the dangers of synthetic media and, in particular, deepfakes.

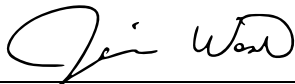
⁴⁴ Vaccari, Cristian, and Andrew Chadwick. "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news." *Social media+ society* 6.1 (2020): 2056305120903408.

23. Protecting the integrity of elections is one of the most important things we can do to preserve democracy. Minnesota’s deepfake law helps achieve this. It is one of the few deterrence mechanisms for reducing the spread of a new and powerful technology that mimics real people and situations through images, videos and audio that are extremely hard and time consuming to parse from reality. Left unchecked, the ease of creation, low cost, scalability potential, and the difficulty in debunking this content could be highly disruptive to an election, especially since the technology is relatively new and much of the population will not be aware of these capabilities or have access to the tools to check the veracity of the content. Some may argue that the ‘market of ideas’—the idea that open discourse and the competition of ideas will lead to truth—can correct and properly rebut depictions of people and events. In the days prior to the internet and social media, in particular, this may have been true. In our current digital world, unfortunately the market of ideas doesn’t work as John Milton and John Stuart Mills may have envisioned nearly 200 years ago when our information environments were much different. Today, information overload, bots and inauthentic account, echo chambers, algorithmic amplification and power imbalances in our digital world create a marketplace where not all voices have equal access. One of the few mechanisms for countering these challenges is the Minnesota deepfake law that can at least add some friction to the production of blatantly false, but highly convincing and potentially market-damaging content.

[Signature on the Following Page]

PURSUANT TO 28 U.S.C. § 1746, I DECLARE UNDER PENALTY OF PERJURY THAT EVERYTHING I HAVE STATED IN THIS DOCUMENT IS TRUE AND CORRECT.

Dated: October 31, 2024



JEVIN WEST